Full wwPDB X-ray Structure Validation Report ⓘ

Feb 11, 2026 – 04:21 PM EST

PDB ID	:	9YZQ / pdb_00009yzq
Title	:	Isoreticular co-crystal 1 with asymmetrical expanded duplex (31mer) containing insert sequence TGATGAGCAG and loaded with Engrailed homeodomain enhanced Green fluorescent protein fusion
Authors	:	Shields, E.T.; Slaughter, C.K.; Magna, E.N.; Snow, C.D.
Deposited on	:	2025-10-30
Resolution	:	3.75 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

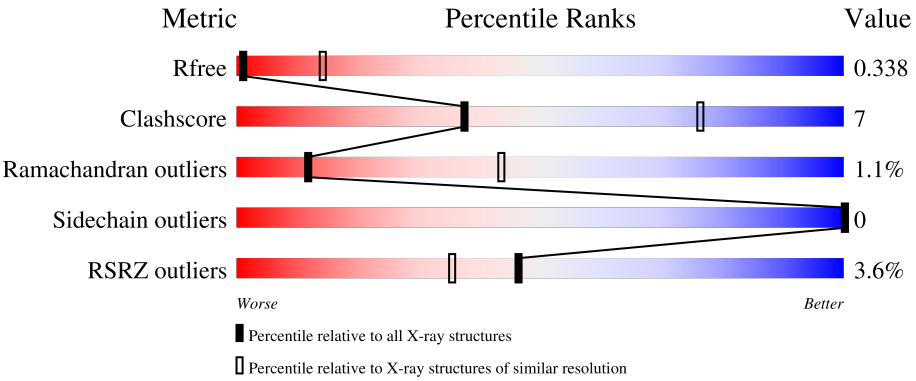
MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.48

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.75 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1256 (3.92-3.60)
Clashscore	180529	1321 (3.92-3.60)
Ramachandran outliers	177936	1293 (3.92-3.60)
Sidechain outliers	177891	1288 (3.92-3.60)
RSRZ outliers	164620	1256 (3.92-3.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	31	45% 55%
2	B	31	58% 42%
3	C	263	3% 73% 8% 18%
4	D	313	% 16% 82%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3198 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	31	633	302	115	185	31	0	0	0

- Molecule 2 is a DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	31	638	304	116	187	31	0	0	0

- Molecule 3 is a protein called Replication initiation protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	215	1614	1014	285	309	6	0	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-11	MET	-	initiating methionine	UNP P03856
C	-10	ARG	-	expression tag	UNP P03856
C	-9	GLY	-	expression tag	UNP P03856
C	-8	SER	-	expression tag	UNP P03856
C	-7	HIS	-	expression tag	UNP P03856
C	-6	HIS	-	expression tag	UNP P03856
C	-5	HIS	-	expression tag	UNP P03856
C	-4	HIS	-	expression tag	UNP P03856
C	-3	HIS	-	expression tag	UNP P03856
C	-2	HIS	-	expression tag	UNP P03856
C	-1	GLY	-	expression tag	UNP P03856
C	0	SER	-	expression tag	UNP P03856
C	118	PRO	ARG	conflict	UNP P03856

- Molecule 4 is a protein called Segmentation polarity homeobox protein engrailed,Green fluorescent protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	D	57	Total	C	N	O	0	0	0
			311	187	60	64			

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	MET	-	initiating methionine	UNP P02836
D	61	THR	-	linker	UNP P02836
D	62	SER	-	linker	UNP P02836
D	63	GLN	-	linker	UNP P02836
D	64	PHE	-	linker	UNP P02836
D	65	TYR	-	linker	UNP P02836
D	66	LEU	-	linker	UNP P02836
D	67	ASN	-	linker	UNP P02836
D	68	GLU	-	linker	UNP P02836
D	70	VAL	-	insertion	UNP P42212
D	133	LEU	PHE	conflict	UNP P42212
D	134	THR	SER	conflict	UNP P42212
D	300	LEU	HIS	conflict	UNP P42212
D	308	HIS	-	expression tag	UNP P42212
D	309	HIS	-	expression tag	UNP P42212
D	310	HIS	-	expression tag	UNP P42212
D	311	HIS	-	expression tag	UNP P42212
D	312	HIS	-	expression tag	UNP P42212
D	313	HIS	-	expression tag	UNP P42212

- Molecule 5 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		
5	C	1	Total	Mg	0	0
			1	1		

3 Residue-property plots (i)

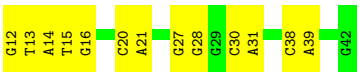
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA (31-MER)

Chain A:

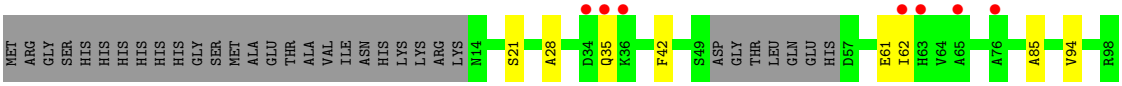
• Molecule 2: DNA (31-MER)

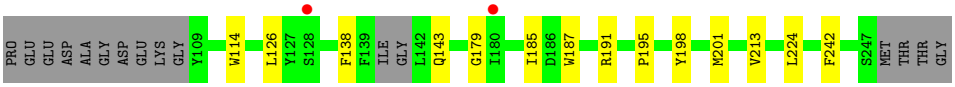
Chain B:



• Molecule 3: Replication initiation protein

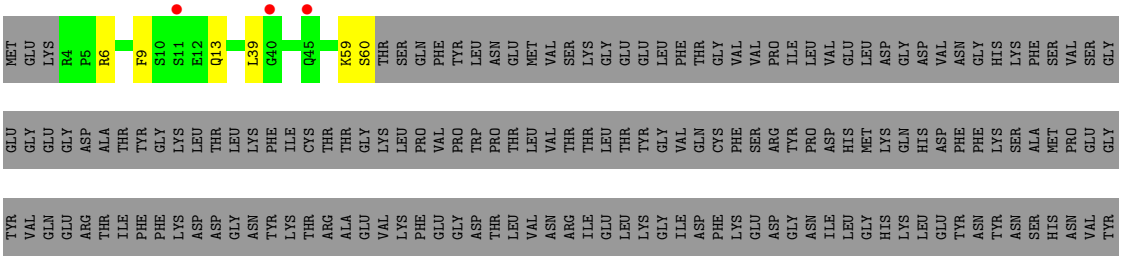
Chain C:





• Molecule 4: Segmentation polarity homeobox protein engrailed,Green fluorescent protein

Chain D:



ILE MET ALA ASP LYS GLN LYS ASN GLY ILE LYS VAL ASN PHE LYS ILE ARG HIS ASN ILE GLU ASP GLY SER VAL GLN LEU ALA ASP HIS TYR GLN GLN ASN THR PRO ILE GLY ASP GLY PRO VAL LEU LEU PRO ASP ASN HIS TYR LEU SER THR GLN SER ALA LEU SER LYS ASP PRO

ASN GLU LYS ARG ASP HIS MET VAL LEU LEU GLU PHE VAL THR ALA ALA GLY ILE THR LEU GLY MET ASP GLU LEU TYR LYS HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	74.12Å 122.40Å 140.44Å 90.00° 90.38° 90.00°	Depositor
Resolution (Å)	46.92 – 3.75 46.92 – 3.75	Depositor EDS
% Data completeness (in resolution range)	97.7 (46.92-3.75) 97.6 (46.92-3.75)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.17 (at 3.77Å)	Xtriage
Refinement program	PHENIX 1.21.2_5419	Depositor
R, R_{free}	0.295 , 0.338 0.295 , 0.338	Depositor DCC
R_{free} test set	1267 reflections (9.77%)	wwPDB-VP
Wilson B-factor (Å^2)	135.0	Xtriage
Anisotropy	0.287	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 160.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.20$	Xtriage
Estimated twinning fraction	0.099 for h,-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3198	wwPDB-VP
Average B, all atoms (Å^2)	221.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.19% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.35	0/709	0.51	0/1091
2	B	0.49	0/715	0.75	0/1102
3	C	0.32	0/1650	0.59	0/2241
4	D	0.49	0/312	0.72	0/432
All	All	0.38	0/3386	0.63	0/4866

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	633	0	350	12	0
2	B	638	0	351	8	0
3	C	1614	0	1450	14	0
4	D	311	0	174	2	0
5	A	1	0	0	0	0
5	C	1	0	0	0	0
All	All	3198	0	2325	36	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 7.

All (36) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:185:ILE:HG21	3:C:201:MET:HE1	1.64	0.80
4:D:59:LYS:O	4:D:60:SER:CB	2.34	0.75
1:A:10:DA:H2"	1:A:11:DA:H5"	1.79	0.65
1:A:3:DT:H2'	1:A:4:DG:C8	2.36	0.60
3:C:35:GLN:HG3	3:C:85:ALA:HB1	1.85	0.59
1:A:9:DA:H2"	1:A:10:DA:C8	2.38	0.57
3:C:187:TRP:CE2	3:C:191:ARG:HG3	2.40	0.57
3:C:61[A]:GLU:HB2	3:C:126:LEU:HD11	1.88	0.55
1:A:8:DC:H2"	1:A:9:DA:H5"	1.90	0.54
2:B:27:DG:H2'	2:B:28:DG:C8	2.42	0.54
2:B:12:DG:H2"	2:B:13:DT:H5"	1.90	0.54
3:C:94:VAL:HB	3:C:114:TRP:HE1	1.73	0.54
3:C:187:TRP:CZ2	3:C:191:ARG:HG3	2.44	0.53
3:C:21:SER:HB3	3:C:138:PHE:O	2.08	0.52
1:A:13:DT:H2'	1:A:14:DG:C8	2.46	0.51
2:B:15:DT:H2"	2:B:16:DG:H5'	1.93	0.50
1:A:25:DA:H2"	1:A:26:DG:C8	2.48	0.49
1:A:2:DC:H2'	1:A:3:DT:C6	2.49	0.48
2:B:30:DC:H2"	2:B:31:DA:N7	2.29	0.47
1:A:23:DT:H2"	1:A:24:DG:C8	2.49	0.47
3:C:213:VAL:HG22	3:C:224:LEU:HD23	1.97	0.47
1:A:8:DC:H2"	1:A:9:DA:C8	2.50	0.47
3:C:61[B]:GLU:HB3	3:C:126:LEU:HD11	1.97	0.46
1:A:30:DT:H2"	1:A:31:DA:C8	2.51	0.46
2:B:27:DG:H2'	2:B:28:DG:H8	1.81	0.45
3:C:179:GLY:HA3	3:C:242:PHE:CZ	2.53	0.44
4:D:13:GLN:HA	4:D:39:LEU:HD22	1.99	0.44
1:A:28:DC:H2"	1:A:29:DA:C8	2.53	0.43
2:B:38:DC:H2"	2:B:39:DA:C8	2.54	0.42
3:C:185:ILE:HG21	3:C:201:MET:CE	2.44	0.42
3:C:42:PHE:CE1	3:C:62:ILE:HD11	2.55	0.42
2:B:14:DA:H2"	2:B:15:DT:C6	2.56	0.41
3:C:195:PRO:HD2	3:C:198:TYR:CD2	2.56	0.41
1:A:8:DC:H2"	1:A:9:DA:H8	1.86	0.41
2:B:20:DC:H2"	2:B:21:DA:C8	2.55	0.41
3:C:28:ALA:HA	3:C:191:ARG:O	2.21	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
3	C	208/263 (79%)	203 (98%)	4 (2%)	1 (0%)	25 58
4	D	55/313 (18%)	51 (93%)	2 (4%)	2 (4%)	3 24
All	All	263/576 (46%)	254 (97%)	6 (2%)	3 (1%)	12 43

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	143	GLN
4	D	6	ARG
4	D	9	PHE

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	C	158/236 (67%)	158 (100%)	0	100 100
4	D	9/278 (3%)	9 (100%)	0	100 100
All	All	167/514 (32%)	167 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no oligosaccharides in this entry.

5.6 Ligand geometry

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	31/31 (100%)	-0.45	0 100 100	215, 233, 312, 328	0
2	B	31/31 (100%)	-0.40	0 100 100	192, 239, 326, 402	0
3	C	215/263 (81%)	-0.14	9 (4%) 41 32	87, 179, 282, 413	1 (0%)
4	D	57/313 (18%)	-0.14	3 (5%) 33 27	131, 256, 433, 572	0
All	All	334/638 (52%)	-0.19	12 (3%) 46 35	87, 211, 326, 572	1 (0%)

All (12) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	45	GLN	7.9
3	C	76	ALA	5.0
3	C	63	HIS	4.9
3	C	62	ILE	4.1
3	C	65	ALA	3.5
3	C	128	SER	2.8
4	D	40	GLY	2.4
3	C	180	ILE	2.2
3	C	34	ASP	2.2
3	C	35	GLN	2.1
3	C	36	LYS	2.1
4	D	11	SER	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	MG	A	101	1/1	0.94	0.06	69,69,69,69	0
5	MG	C	301	1/1	0.97	0.03	29,29,29,29	0

6.5 Other polymers

There are no such residues in this entry.